A Comparison of System Latency in Copper and Optical PCIe® Links

Kevin Burt

Technical Marketing

Samtec

Disclaimer

<u>Presentation Disclaimer</u>: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Acknowledgments

Hugo Kohmann **Dolphin Interconnect Solutions**

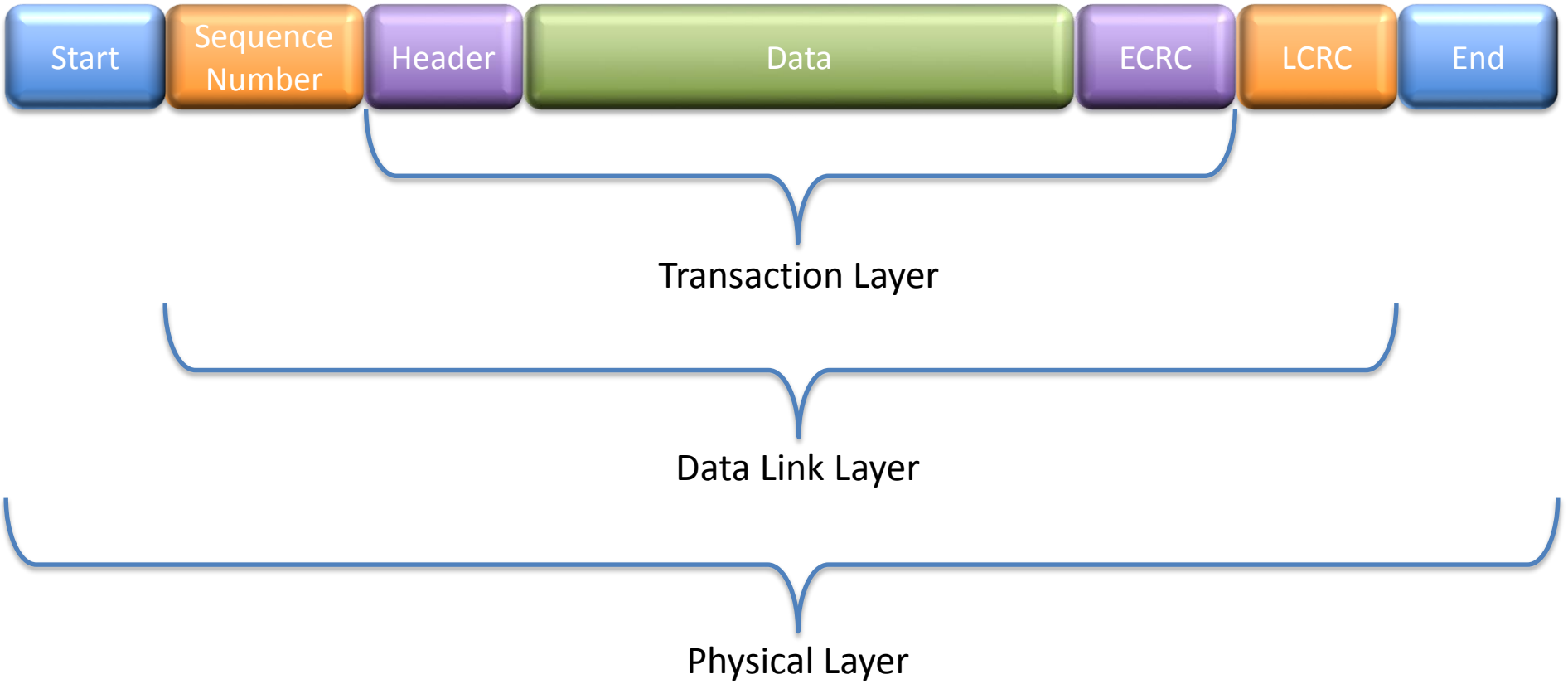
Preben N. Olsen **Dolphin Interconnect Solutions**

Agenda

- **Latency Components**
  - Physical layer
  - Link Layer
  - Transaction Layer
  - Operating System
- **Measurement Results**
  - Copper vs. Optical
  - Linux and Real Time Operating System
- **Summary**

Latency Components

PCIe® Layers



Latency Components

Physical Layer

Physics – Propagation Velocity

- **Velocity or Propagation (V_p) of a wave in a medium is given by:**

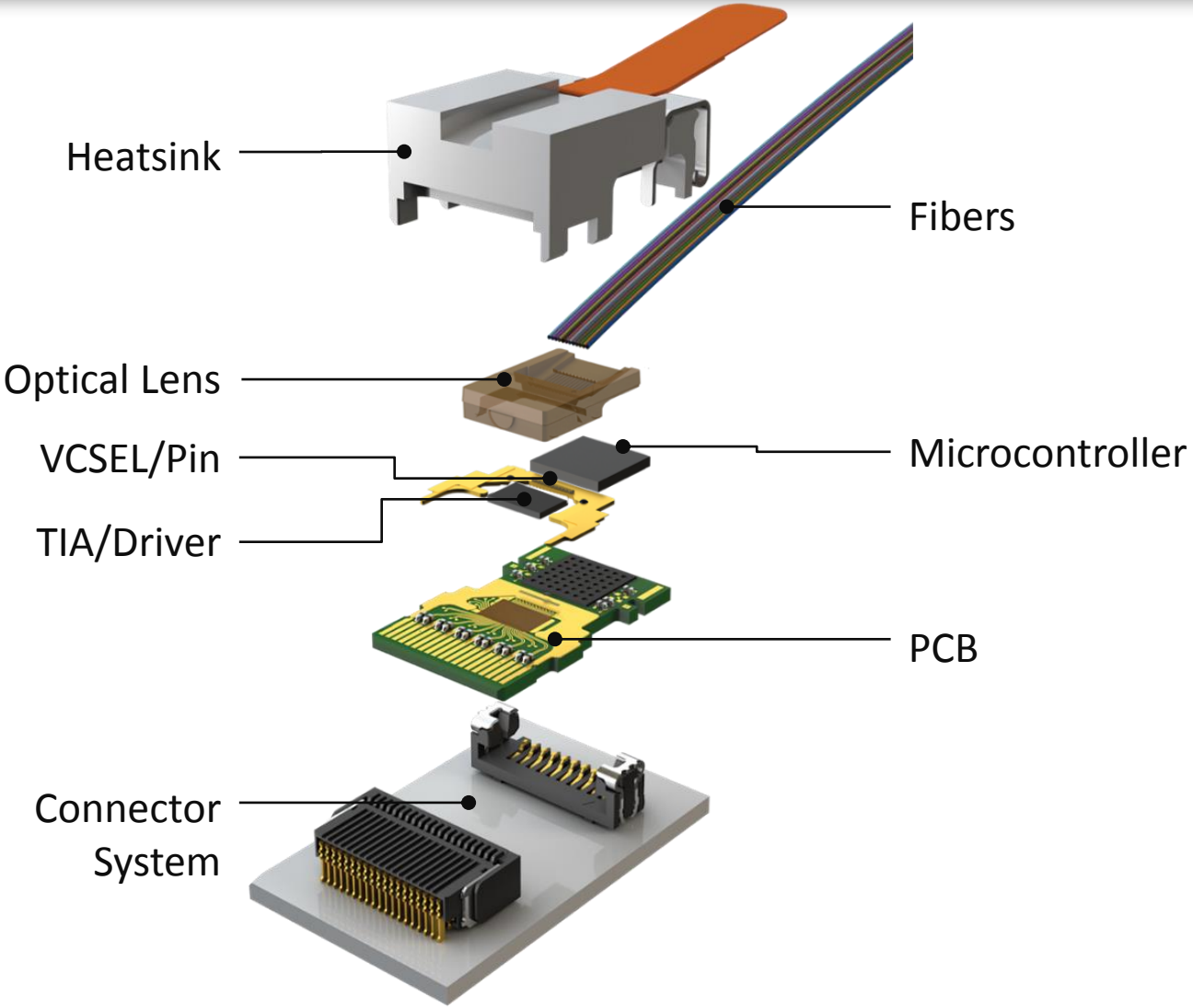
$$V_p = \frac{c}{\sqrt{\kappa}}$$

Where:

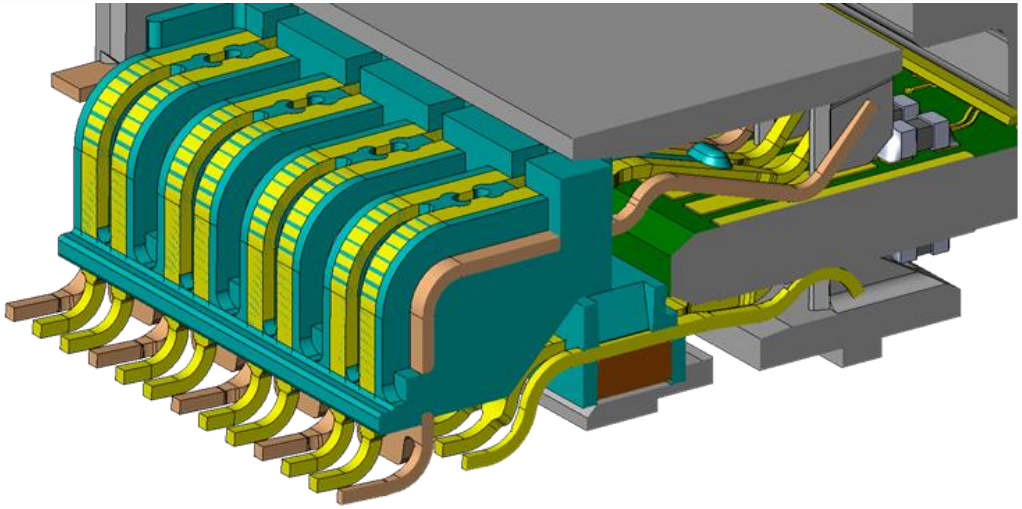
c = speed of light in a vacuum (299,792,458 m/s)

κ = dielectric constant of the medium

Anatomy of an Optical Interconnect



Connector System



Connector has a short row and a long row

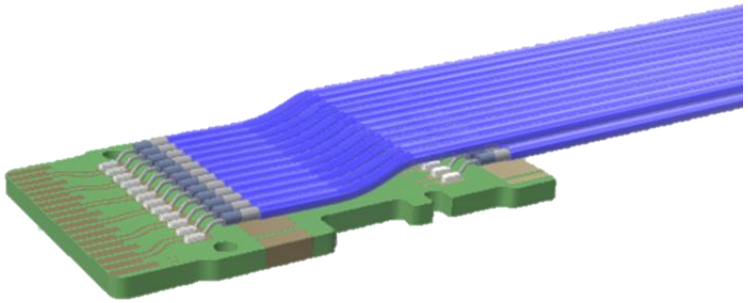
Full channel uses both, with one on each end

- Long row = 7.5 mm
- Short row = 5.7 mm

Propagation speed = 6.3 ps/mm

Total latency = 83 ps

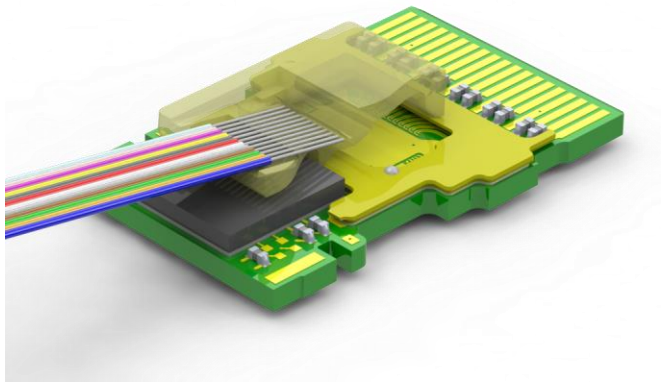
Copper PCB



- **PCB propagation speed = 6.24 ps/mm**
  - Short channel = 5.2 mm
  - Long channel = 11.2 mm
- **Again, full link uses both**

- **Total latency = 102 ps**

Optical PCB



- **PCB propagation speed = 6.24 ps/mm**
  - length = 6.36 mm
- **PCB latency = 61 ps**
- **ICs**
  - Feature and vendor specific
  - 100 – 500 ps per chip

Copper / Fiber

- **Nothing is faster than the speed of light** in a vacuum
  - Propagation speed in fiber is 5.13 ns/m
  - Propagation speed in copper is 4.79 ns/m
- **Latency is dominated by length, *l***
  - Copper = 0.18 + 4.8*l*
  - Optical = 1.1 + 5.1*l*
- **Optical can do 100 m at 8 GT/s…**
  - Significant increase to “time of flight”

Latency Components

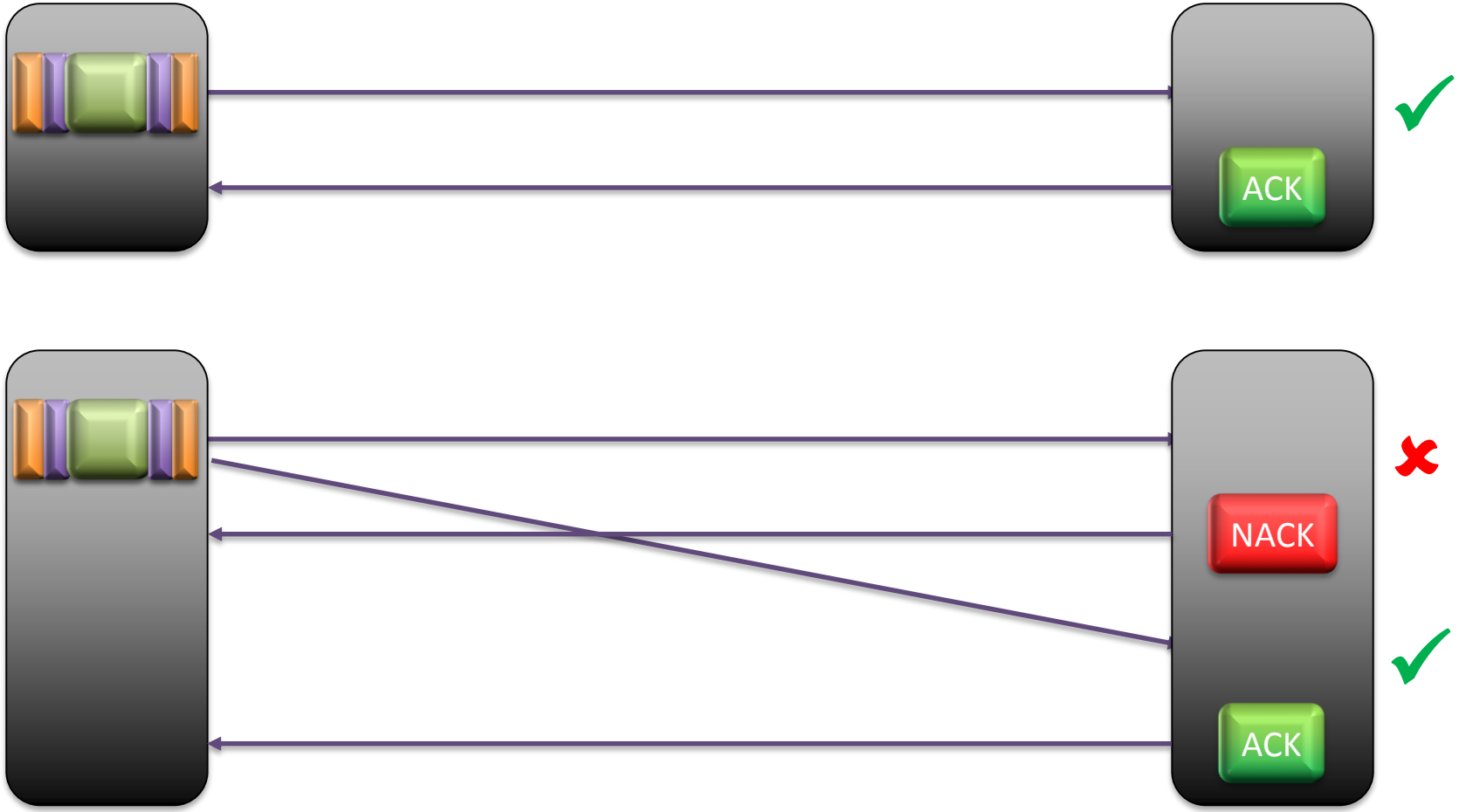
Link Layer
Transaction Layer

Robust Transmission

- **Key feature of PCIe is the inherent robustness of the transmission**
  - Sequence Number
  - CRC

  - ACK / NACK
  - Flow Control

ACK / NACK



Sequence Flow

Latency for the NACKed data is > 3x that for a good packet

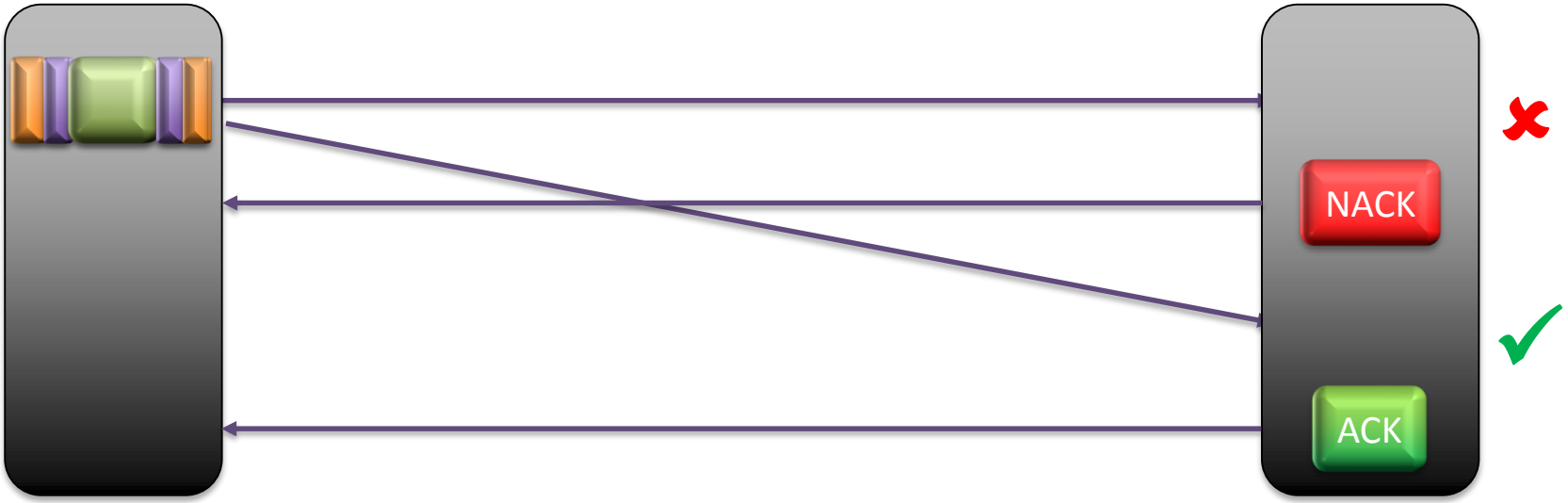
However packet ordering is also preserved

All subsequent TLP are discarded until the replayed DLPs are detected

ACK / NACK Replay Timer

- **Transmitter cannot assume a transaction has been correctly received**
  - Has to have an ACK

- **If an ACK is not received in an appropriate amount of time, the transmitter has to resend all unacknowledged TLPs**



Buffer Size and Flow Control

PCIe control devices have to balance buffer size and cost

- **Too little results in increased latency**
- **Adding transistors adds cost**

Flow Control prevents data loss through credits

Flow Control

- During initialization, each side reports it's buffer size
- The transmitter can only send enough data to fill the buffer
- It then needs to wait until there is an update on how many have been processed
- Only then can it send more data

- Data become bursty and latency increases

Latency Components

Operating System

Operating System

- **An application is allocated to a CPU and will run uninterrupted until it completes its tasks**
  - More often regular operating systems will re-schedule the application or interrupt is when the OS issues an internal inter-CPU interrupt

- **This is hardly noticeable in regular applications, but in latency critical systems timing requirements may be violated**
  - Solution is to use a Real-Time Operating System that provides CPU shielding
  - Prevents interrupts

Measurements

Method

- **Used Dolphin's SISCI API ping pong benchmark to measure best, average and worst case latency for a sequence of ping pong data transfers**

- **Latency is for a half-way round trip between two systems**
  - High precision timer used to measure the round-trip latency for each transfer

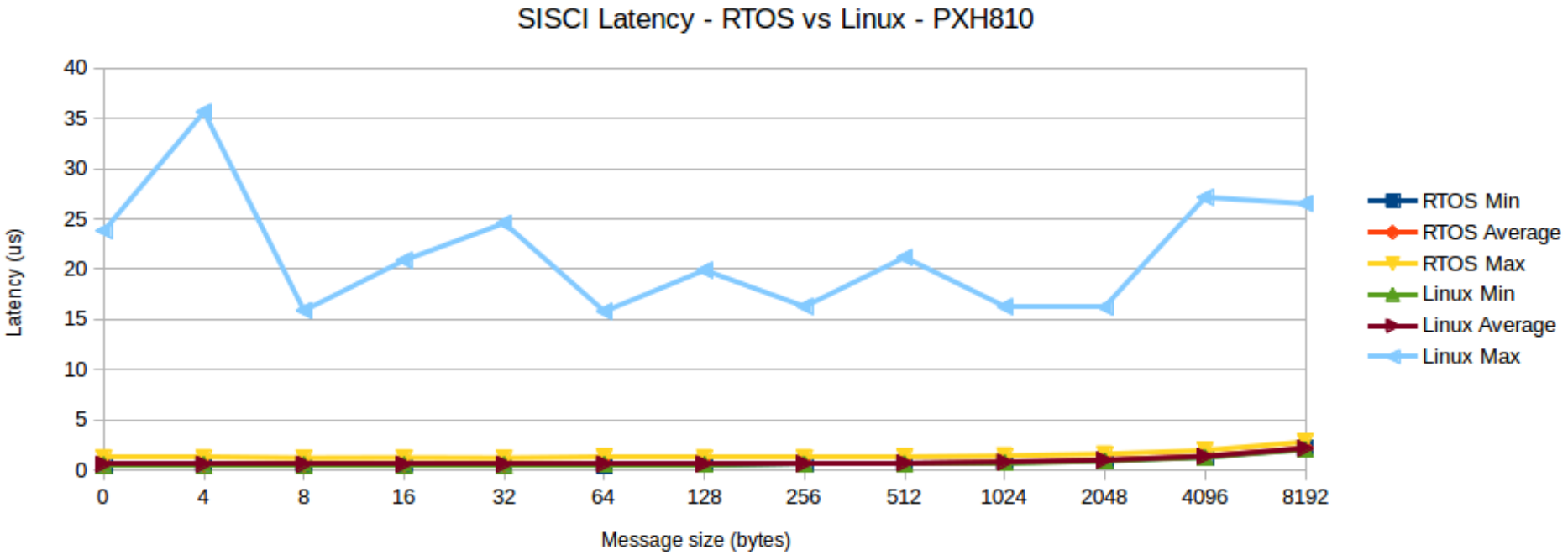
Lab Setup PCIe 3.0 x8

- **Standard servers**
- **Dolphin PXH810 cards**
  - PCIe 3.0 x8, iPass, NTB chipset
- **Dolphin eXpressWare 5.5.0**
- **Linux and a RTOS**
- **PCIe 2.0 0.5 m copper**
- **Samtec iPass 10 and 100 m fiber optics**

Lab Setup PCIe 3.0 x16

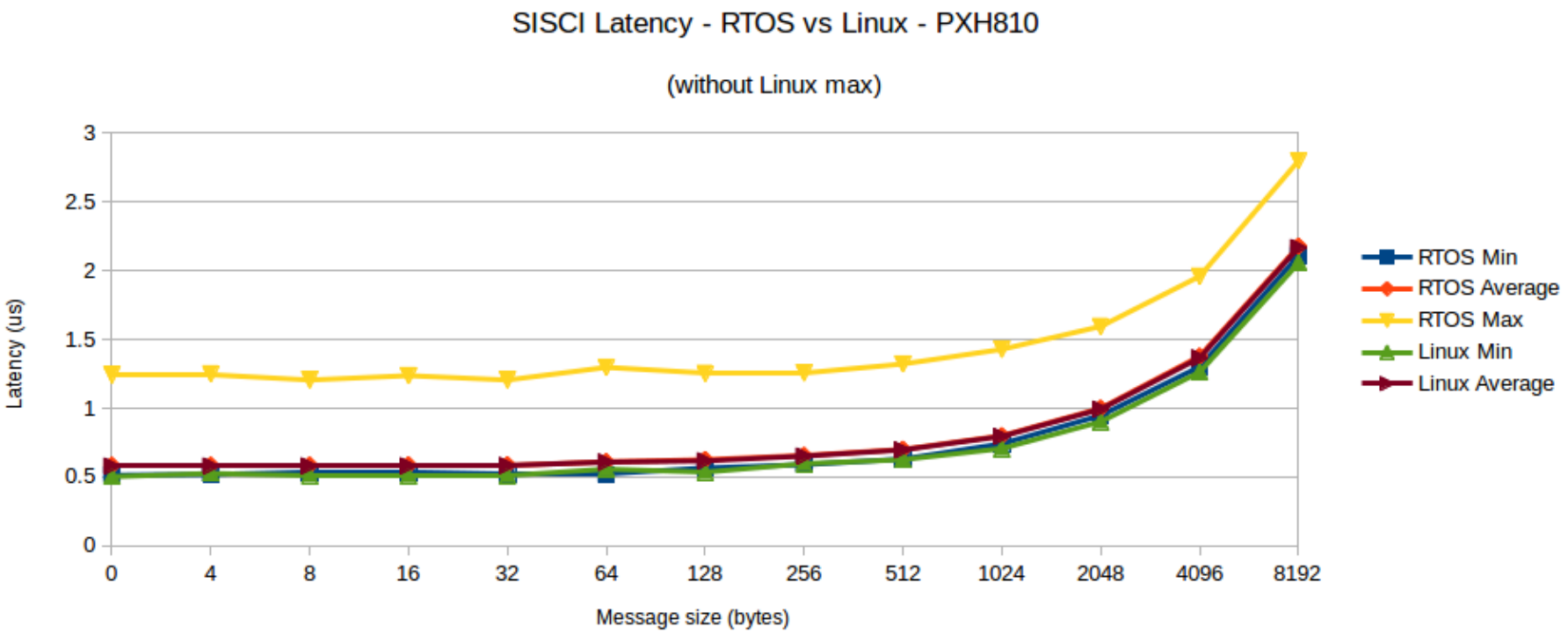
- **Standard servers**
- **Linux**
- **Dolphin PXH830 cards**
  - PCIe 3.0 x8, SFF-8644, NTB chipset
- **Samtec / Dolphin PCOA cards**
  - PCIe 3.0 x16, FireFly
- **Dolphin eXpressWare 5.5.0**
- **MiniSAS-HD cable 0.5 m copper**
- **Samtec FireFly, 10, 50 and 100 m fiber optics**

Comparing OS



- **Minimum = 510 ns**
- **Average = 540 ns**
- **Worst Case = 1.24 µs (RTOS)**
  35 µs (Linux)

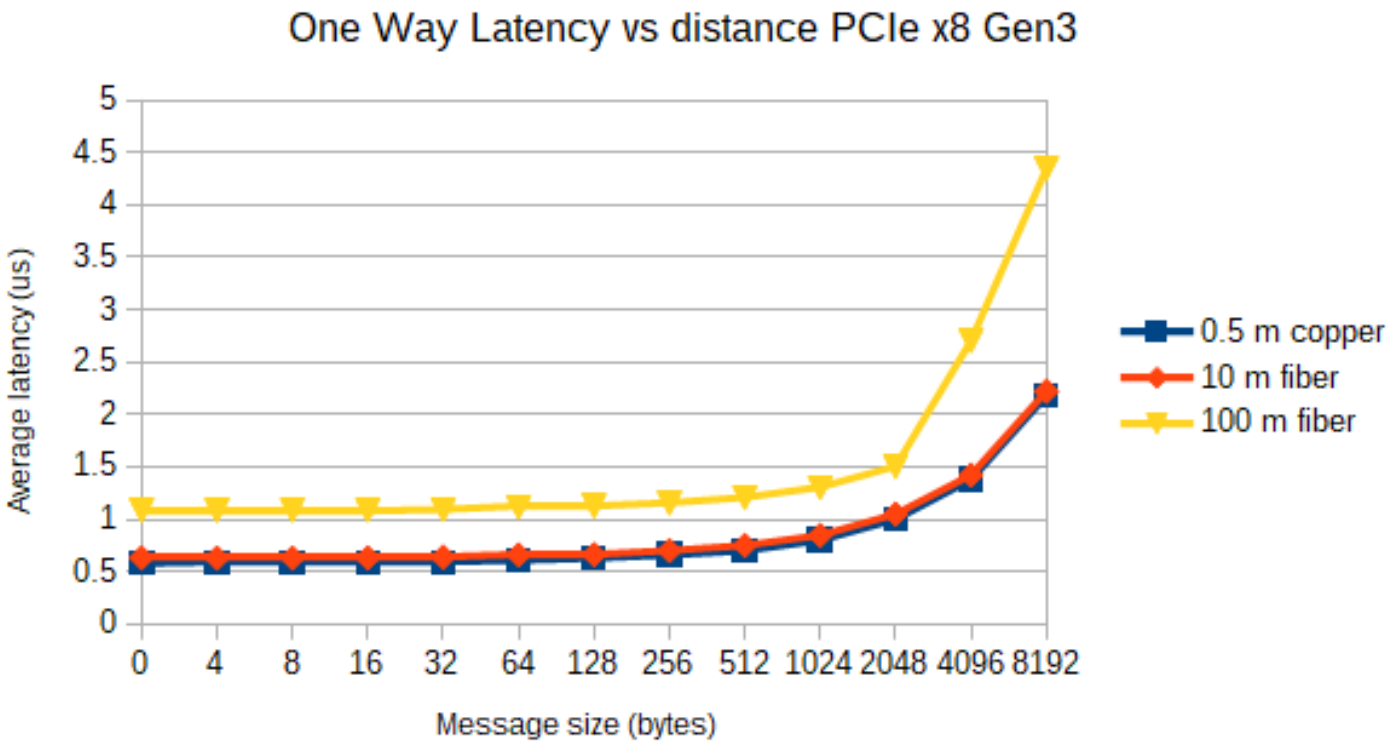
Same Results as Previous Without Linux



0.5 meter copper cable

RTOS worst case latency is very close to average

Comparing distance



4K and 8K message latency, 100 meter fiber

latency increases due to out of PCIe credits

Summary

Summary

Latency is a key benefit of optical systems

- **As links get longer time of flight becomes important**
  - Propagation delay
  - Protocol related resends and pauses
- **Copper / Optical latency generally follows the theory**
  - Latency disproportionally increases at long lengths and larger packet size due to flow control

A Real Time Operating System, can help reduce the inherent latency caused by system interrupts

Thank you for attending the PCI-SIG Developers Conference Asia-Pacific Tour 2017.

For more information please go to www.pcisig.com